

University of Groningen

## Secondary Resources in the Bio-Based Economy

Davis, Chris B.; Aid, Graham; Zhu, Ben

*Published in:*  
Waste and Biomass Valorization

*DOI:*  
[10.1007/s12649-017-9975-0](https://doi.org/10.1007/s12649-017-9975-0)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Davis, C. B., Aid, G., & Zhu, B. (2017). Secondary Resources in the Bio-Based Economy: A Computer Assisted Survey of Value Pathways in Academic Literature. *Waste and Biomass Valorization*, 8(7), 2229-2246. <https://doi.org/10.1007/s12649-017-9975-0>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Secondary Resources in the Bio-Based Economy: A Computer Assisted Survey of Value Pathways in Academic Literature

Chris B. Davis<sup>1</sup>  · Graham Aid<sup>2,3</sup> · Ben Zhu<sup>4</sup>

Received: 14 December 2016 / Accepted: 22 May 2017 / Published online: 2 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Research on value pathways for organic wastes has been steadily increasing in recent decades. There have been few considerably broad overview studies of such materials and their valuation potential in the bio-based economy in part because of the vast multitude of materials and processes that can be used to produce energy carriers, chemicals, and materials of value. This article explores how automated data analysis approaches can help in analyzing large bodies of text to distill and present potential value pathways for secondary (waste) bio-based materials. The study employed multiple methods (literature collection, topic modelling, and co-occurrence analysis) on a collection of abstracts from 53,292 academic articles covering technologies, applications, and products (TAPs) for bio-based wastes. The results of both the topic modelling and co-occurrence analysis are presented as online interactive web pages. The topic modelling presented an overview of research clusters related to secondary organic resources, processes, and disciplines. The co-occurrence analysis helped to understand which TAPs are researched in relation

to a broad spectrum of organic wastes. Co-occurrences were evaluated using the Normalized Pointwise Mutual Information measure to locate terms which co-occur more frequently than would be expected by chance. Through the use of detailed lists of organic wastes and TAPs, the co-occurrence method mapped out 7118 unique intersections between 473 specific wastes and 228 TAPs. This technique enables us to find seemingly non-obvious valorization pathways such as the re-use of oyster shells as catalysts for bio-diesel production and bioplastic production from brewery waste. While a proof-of-concept, this work points the way for using Big Data to suggest novel pathways for implementing the Circular Economy.

**Keywords** By-product · Waste valorization · Circular economy · Recycling · Industrial symbiosis · Big Data

## Introduction

Many regions and nations see the development of the bio-based economy as a strategic step in dealing with rapid depletion of many (especially fossil based) resources, increasing environmental pressures and climate change [1–3]. Understanding potential bio-based economy developments will be key in furthering implementation and innovation in parallel with sustainability strategies. Traditionally, the bio-based economy includes primary production industries, such as forestry, fisheries, aquaculture, agriculture, and industries utilizing biological resources, such as paper and food industries. Increasingly the pharmaceutical and chemical industries have been integrated into the bio-based economy through the increased use of bio-based raw materials. Additionally, the waste management industry has been building up its presence in the bio-based

**Electronic supplementary material** The online version of this article (doi:10.1007/s12649-017-9975-0) contains supplementary material, which is available to authorized users.

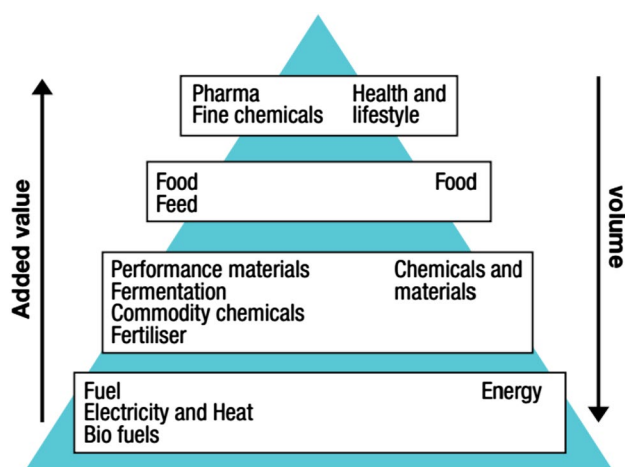
✉ Chris B. Davis  
c.b.davis@rug.nl

<sup>1</sup> Center for Energy and Environmental Sciences, University of Groningen, Groningen, The Netherlands

<sup>2</sup> Department of Research and Development, Ragn-Sells AB, Stockholm, Sweden

<sup>3</sup> Division for Environmental Technology and Management, Linköping University, Linköping, Sweden

<sup>4</sup> Department of Engineering Systems and Services, TU Delft, Delft, The Netherlands



**Fig. 1** The value pyramid of the bio-based economy as used by Green European Foundation [4] and developed by BioBased Economy Netherlands [5]

economy via valuation of secondary (waste) resources such as food waste, agricultural waste, nutrient rich sludge, and ashes.

As shown in Fig. 1, products from the bio-based economy range in value and market size. For example, pharmaceuticals and fine chemicals tend to have high value with lower market sizes, while food, feed, performance materials, and fertilizers tend to have lower values with relatively larger market sizes. Finally, the production of biofuels, electricity, and heat normally has the lowest per-weight value in relation to other bio-based economy products. In order to strategically grow valuable and competitive bio-based economies, system solutions for innovatively cascading and cycling material and value via several interconnected processes will be required. Such innovative systems solutions can be supported by, among other things, making use of dense knowledge in regards to bio-based material streams and potential processing pathways.

## Aim and Objectives

How could someone (a researcher, a waste management developer, a market analyst, etc.) entering the field of the bio-based economy begin to assess the main topics, emerging technologies, or portfolio of potentially undervalued materials in this especially broad topic? In this article, we see our main problem owner as being a waste management development professional who has to “matchmake” between a large number of diverse bio-wastes and potential value pathways. Specifically, *this is a problem of reducing a large permutation space of potential combinations of wastes and value pathways, without excluding novel unexpected combinations that may lead to new valorization*

*opportunities*. By some estimates<sup>1</sup>, the amount of potentially relevant literature is doubling roughly every 5 years, and assuming that this trend continues, it will become more difficult to effectively process this information without some sort of automated assistance.

There are a few meta-studies of valuation processes available to secondary organic resources [6–9]. While these studies are quite useful in their detail, they are far from comprehensive, often focusing on certain types of secondary resources such as food waste, sludge and effluents, agricultural wastes, etc. In the waste management sector much of the development of processes for bio-materials has focused on robust (take all) processes which create commodities on the low value end of bio-based product scale—such as fertilizers via composting [10], fuels via digestion [11], and heat and energy through combustion.

This article aims to test automated methods for analyzing the increasing amount of knowledge being communicated through scientific literature related to valuation pathways for organic wastes. In line with these aims, this study specifically looks to the following objectives:

1. Describe, apply and evaluate specific methods for providing an overview of large amounts of scientific literature in the subject area of interest,
2. Distinguish prominent topics within waste and resource management of organic materials,
3. Present a broad range of valuation pathways for organic waste material,
4. Make non-copyrighted data openly available for utilization and further development.

## Methods

As the analysis involved a very large volume of text, several methods were employed to be able to summarize the text and highlight patterns of potential interest. In analyzing the literature, we aimed to gain a high level overview of macro-trends, while also being able to pinpoint micro-trends that may not be widespread yet, but illustrate interesting directions with regard to how organic waste is processed. In general, we are interested in identifying the general topics being discussed and then performing a more fine-grained analysis about the entities involved. In particular, we were looking for interesting combinations of entities (e.g. “oyster shells” as a catalyst for “biodiesel” production) which may indicate new pathways being explored in the bio-economy.

<sup>1</sup> Based on our literature search, see Fig. 5.

## Literature Collection

The literature was collected via search terms that aimed to locate articles describing processing of organic waste from the years 1995 to 2014. The queries, listed in Appendix A.1, were run on both Scopus and Web of Science (WOS), resulting in 53,292 distinct matching articles. While it would be ideal to analyze the full text of these articles, all of our analysis is restricted to article titles, keywords and abstracts, due to limitations on automated full text downloads for Scopus and Web of Science. As documented more in the “Results”, there is overlap in coverage between Scopus and Web of Science, we had to remove duplicate articles. We performed this step using OpenRefine<sup>2</sup> and the “cluster and edit” feature which uses several string distance metrics to group together text strings that have similar forms.

## Topic Modelling

This study employed topic modelling [12] as both a way to refine the literature search terms as well as to address the objective of “distinguishing prominent topics” in the collected literature. Topic modelling is a means of identifying the dominant themes that occur across a collection of documents. The topics are not predetermined by a user, but rather are based on statistical patterns of words that commonly co-occur. The advantage of this is that it allows for grouping the abstracts together based on their contents instead of the biases of those examining the text. For example, if one were to ask a group of researchers to read the same set of literature and describe what they thought were the dominant topics, their answers would likely differ based on their own interests and background knowledge.

A key value of topic modelling is that it aids in providing a map of what literature has been collected. In early stages of analysis, we noticed a topic involving “organic Rankine cycle” and “waste heat”, which was not relevant for the analysis (as these terms describe a thermodynamic cycle that is suitable for low-grade waste heat). However, we later realized that our search terms at the time were matching these words exactly as we were looking for terms such as “organic” and “waste”. This process of topic modelling the search results allowed us to improve our collection of abstracts by quickly scanning to see if the topics we expected to see were included, and refine our search terms when topics undesirable for further analysis were identified.

The number of topics is specified by the user, and conceptually this is similar to k-means clustering [13], where users specify how many clusters to find, and while a

certain cluster does not necessarily have a specific meaning, objects within that cluster will statistically have a degree of similarity. In general, specifying a small number of topics would indicate general clusters of themes, while a large number of topics would likely distinguish between sub-themes such as ethanol from maize and cellulosic ethanol.

The particular algorithm used for this analysis was Latent Dirichlet Allocation (LDA) [12]. We used the software implementation of this algorithm provided by the MACHine Learning for Language Toolkit (MALLET) [14]. There are several outputs of the software which allow us to characterize the literature collected. First, for every document there is a vector with 50 elements, each corresponding to a topic number, and indicating the document’s relevance within each of them. By examining the elements of this vector, one can see the degrees to which a document may actually be discussing multiple topics. Secondly, for every topic, there is a vector of words and their weights within that topic. These word weights are used when characterizing a document’s relevance to a particular topic. For example, a document may contain a single heavily weighted word, or a combination of lower weighted words.

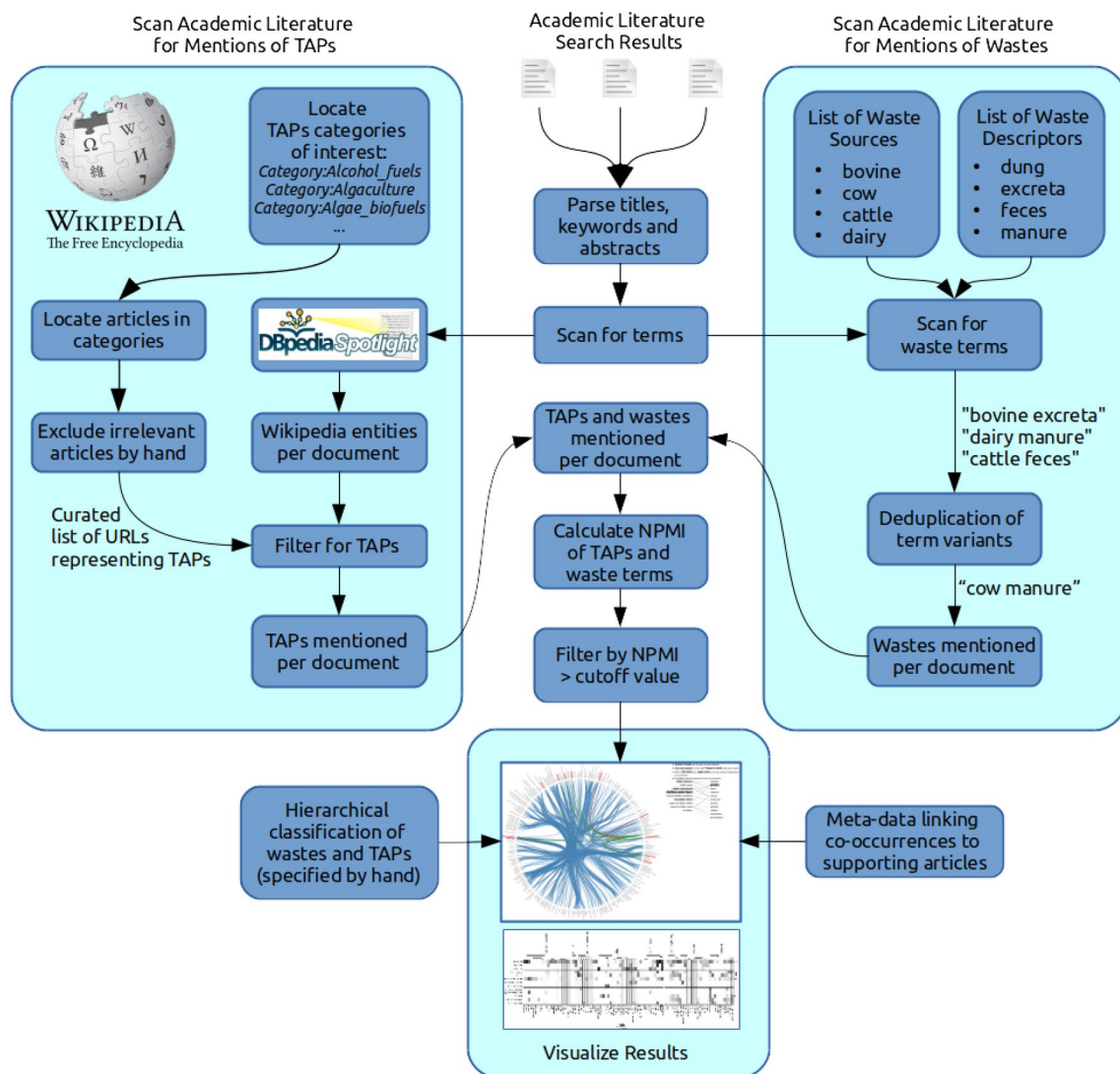
The output of MALLET is a set of matrices which are not easy to interpret on their own. To expose the results in a more user-friendly way, we visualize them with the dfr-browser library<sup>3</sup> for R. The dfr-browser library creates an interactive online visualization that allows one to pivot around different analysis results of the topics, documents and words in a way that helps to show the statistical relations between them. For example, one can see an overview of topics showing their top-weighted words. Per topic, one can find the most relevant documents sorted by score. Per word, one can also see which topics it is prevalent in, which can indicate different contexts for the word. For example, “soil” may occur in topics mentioning actual soil studies, while also appearing in other topics where it is more incidental, such as plant growth.

## Co-occurrence Analysis of Terms

To examine the content of the abstracts in more detail, and to address our third objective to “Identify a broad range of processes utilized for secondary waste material valuation in academic literature”, co-occurrence analysis of entities was performed as illustrated in Fig. 2. Put simply, this process analyzed which types of wastes were mentioned in the literature along with certain types of technologies, applications or final products (TAPs). To achieve this, three steps were required: (a) create a robust list of TAPs, (b) create

<sup>2</sup> <http://openrefine.org/>.

<sup>3</sup> <https://github.com/agoldst/dfr-browser>.



**Fig. 2** The process of the co-occurrence analysis performed in this study. The academic literature sources to be analyzed are read in at the top. The *left* and *right* side show the different approaches used to scan for mentions of wastes and TAPs in that literature. At

the bottom, the co-occurrence analysis results for wastes and TAPs found in academic literature are visualized using different techniques described later in this article

a sufficiently detailed list of wastes, (c) scan academic literature, recording which wastes and TAPs are mentioned in which documents, and (d) analyze and present statistics on the co-occurrence of terms from these lists in the collected literature.

A motivation for this approach is that it is not unusual for researchers to search academic databases for co-occurrences of wastes and TAPs in order to find literature showing how a particular waste can be utilized (e.g. “fish fat” and “biodiesel”). Our goal in co-occurrence analysis is to offer an alternative to performing one search at a time by hand. Instead, using the co-occurrence analysis approach we can scan a large body of literature for a variety of wastes and TAPs, and perform statistical analysis to

highlight combinations which could be useful for researchers or practitioners seeking options for processing wastes.

We use two different strategies to scan academic literature for mentions of TAPs and wastes. For these tasks in general, we needed to have a list of terms to match, a way to match variants of those terms (e.g. plural and singular forms, synonyms, differing adjectives, etc.), and finally a way to match term variants into a single “preferred” version useful for later analysis. The literature we collected contains over 16 million words comprising the titles, keywords and abstracts. A system for term location which is as automated as possible was desired, although as described below, this is not completely possible in particular for the waste terms.



We should emphasize that the lists of wastes and TAPs which we used to find terms of interest are not meant to be exhaustive or authoritative, but rather comprehensive and broad enough to demonstrate the value of our approach. The system we have can automatically regenerate the analysis once better expert-curated lists are created.

### Filtering for TAPs

Creating a list to capture TAPs of interest was difficult as there is not a standard classification<sup>4</sup> that we can use, especially one that can be easily linked to the entities that have been extracted. For this analysis, as shown in Appendix A.3, we compiled a list of 85 Wikipedia categories covering a range of topics in the bio-economy, then created a list of all the article titles within those categories (779 total), and finally filtered this by hand down to a list of 112 article titles which represented an inventory of energy, products, chemicals, processes, and agricultural terms. Furthermore, we amended the list by incorporating a literature review of chemicals mentioned as being promising to focus on within the bio-economy [15].

There are three reasons for using titles of Wikipedia articles to compile this list. First, it gave us a basis to build a sufficiently comprehensive list to demonstrate our approach. Secondly, Wikipedia article URLs (titles) essentially function as unique identifiers, as every distinct concept should only have one Wikipedia article present for it. To be clear, in our analysis we are not using any of the actual *content* of Wikipedia articles, but rather simply the *titles* themselves. The only content we analyze is sourced from academic literature. While people are naturally skeptical of the quality of Wikipedia, one can argue that it is at least extremely comprehensive, with currently 5.4 million articles as of the time of writing. Finally, this approach allowed us to use the DBpedia Spotlight service [16] to automate the process of locating terms of interest found in the abstracts. A more detailed discussion of using DBpedia Spotlight for this purpose is given in Appendix A.4.

Through the DBpedia Spotlight analysis, we can determine for every academic article abstract analyzed, which terms are present which have corresponding Wikipedia titles. Therefore, in an efficient manner we get a very broad scan that allows us to know if an abstract contains terms related to chemicals, organisms, place names, etc. For our analysis we needed to further filter the set of Wikipedia articles found per abstract, and only keep those mentioned in the list of TAPs compiled for the study.

<sup>4</sup> Such as the Harmonized System (HS) or Classification of Products by Activity (CPA) product codes.

### Filtering for Wastes

For locating detailed waste terms in the abstracts, DBpedia Spotlight is not suitable, as Wikipedia does not have article titles at the level of detail that we would like to analyze. Existing waste classifications such as the European List of Waste (LoW)<sup>5</sup> are not suitable either as they are too aggregated.

To enable scanning for detailed waste terms, we first created two lists: one for “waste sources”, and one for “waste descriptors”. As shown in Appendix A.3, the “waste sources” list contains 189 items or locations from which the wastes are produced (e.g. apple, aquaculture, banana, etc.). The “waste descriptors” list contains 156 various forms of waste (e.g. ash, pomace, peel, etc.). These two lists were initially populated using the biomass waste terms mentioned in the ECN Phyllis database [17] and were expanded to include synonyms and other wastes or sources of interest. In scanning for waste terms, we looked for series of terms that contained matches from both lists (e.g. apple + peel, core, pomace, etc.). After locating the terms in the abstracts, we also had to do extensive de-duplication in order to standardize the waste terms as much as possible for later analysis. For example, through this strategy, we were able to successfully locate the terms cow manure, cow excreta, cattle dung, bovine manure and dairy cow feces in the abstracts, but we needed to take an extra step to group these variants into a single “preferred” term.

It is possible that the final list of wastes documented in Appendix A.3.3, may be one of the most extensive lists of organic waste compiled. To compare, the European Waste Catalog (EWC) lists a total of 90 wastes within its 02, 03, and 04 categories<sup>6</sup>, which cover agriculture, forestry, food processing, wood processing, and the textile industries among many others. Data analysis utilizing such lists can be problematic as the waste entities in the EWC and similar lists are often very vague such as:

- 02 06 01—Materials unsuitable for consumption or processing,
- 02 01 03—Plant-tissue waste,
- 02 01 01—Sludges from washing and cleaning,
- 03 03 01—Waste bark and wood,
- 04 02 10—Organic matter from natural products (for example grease, wax),
- 04 02 99—Wastes not otherwise specified.

<sup>5</sup> <http://ec.europa.eu/environment/waste/framework/list.htm>.

<sup>6</sup> 02—wastes from agriculture, horticulture, aquaculture, forestry, hunting and fishing, food preparation and processing; 03—wastes from wood processing and the production of panels and furniture, pulp, paper and cardboard; 04—wastes from the leather, fur and textile industries.

### Co-occurrence Analysis and Visualization

One strategy for analyzing co-occurring terms would be to count how often two terms appear together in all documents. A drawback of this approach is that it would likely only highlight common knowledge or combinations that are already well known and appear in a large number of documents. To get around this issue, we used a different strategy, which was to calculate the Normalized Pointwise Mutual Information (NPMI) of co-occurring terms [18]. The Pointwise Mutual Information is a statistical measure that evaluates whether two terms co-occur more often than would be expected by chance, which thus indicates that there may be some special relation between them. As seen in Eq. 1, it is the ratio of the probability that two terms occur together in a document, divided by the probability that they occur in a document (independent of whether the other is present). The probabilities are based on actual observations. In other words, a value of  $p(x, y) = 0.1$  means that the terms  $x$  and  $y$  are observed to co-occur together in 10% of all the documents collected. The NPMI (Eq. 2) was used to then normalize the PMI scores to between  $-1$  and  $1$ .

$$pmi(x;y) = \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

Pointwise Mutual Information (PMI),

$$npmi(x;y) = \frac{pmi(x;y)}{-\log p(x, y)} \quad (2)$$

Normalized Pointwise Mutual Information (NPMI) as derived from PMI.

After scanning for terms in all the abstracts, observations on the occurrence of TAPs and wastes per abstract were then used to compute NPMI scores for each co-occurrence of a waste and a TAP. A NPMI value of  $1$  means that the terms always co-occur,  $0$  means that they are independent of each other, and  $-1$  means that terms never co-occur.

As with the topic modelling, this step generated a large amount of data which in its raw form was not easily interpretable. To help with this, we created an interactive visualization, which can be viewed at <https://is.gd/wastecircle>, that employed the following strategies. First, wastes and TAPs were arranged in a circle with connecting lines showing co-occurrences that have a NPMI value above the threshold. As detailed in the results, in the literature we were able to locate over 473 specific wastes. To prevent the visualization from becoming too cluttered, we grouped the wastes by their waste sources (e.g. “apple pomace” and “apple peel” show up as “apple”). When someone places their mouse over a term like “apple”, a secondary visualization is shown to give a more detailed view of the specific waste terms and their linkages with TAPs. Then the user

can highlight a combination of a specific waste term and a TAP to see links to the actual literature sources mentioning that combination. This helps users to verify for themselves the exact nature of this co-occurrences and if it is interesting for their purposes.

To further unclutter the circle visualization, we grouped together similar wastes and TAPs along the circumference of the circle, given the categories shown in Appendices A.3.3 and A.2.2. For the TAPs, the categories largely correspond to those found in a literature review by [15], with categories for terms not in that review filled in by hand.

The final circle visualization employed hierarchical edge bundling [19] using an implementation created with the d3.js JavaScript library<sup>7</sup>, which we extended to include the more detailed network visualization and to also show the links to literature mentioning the co-occurrences.

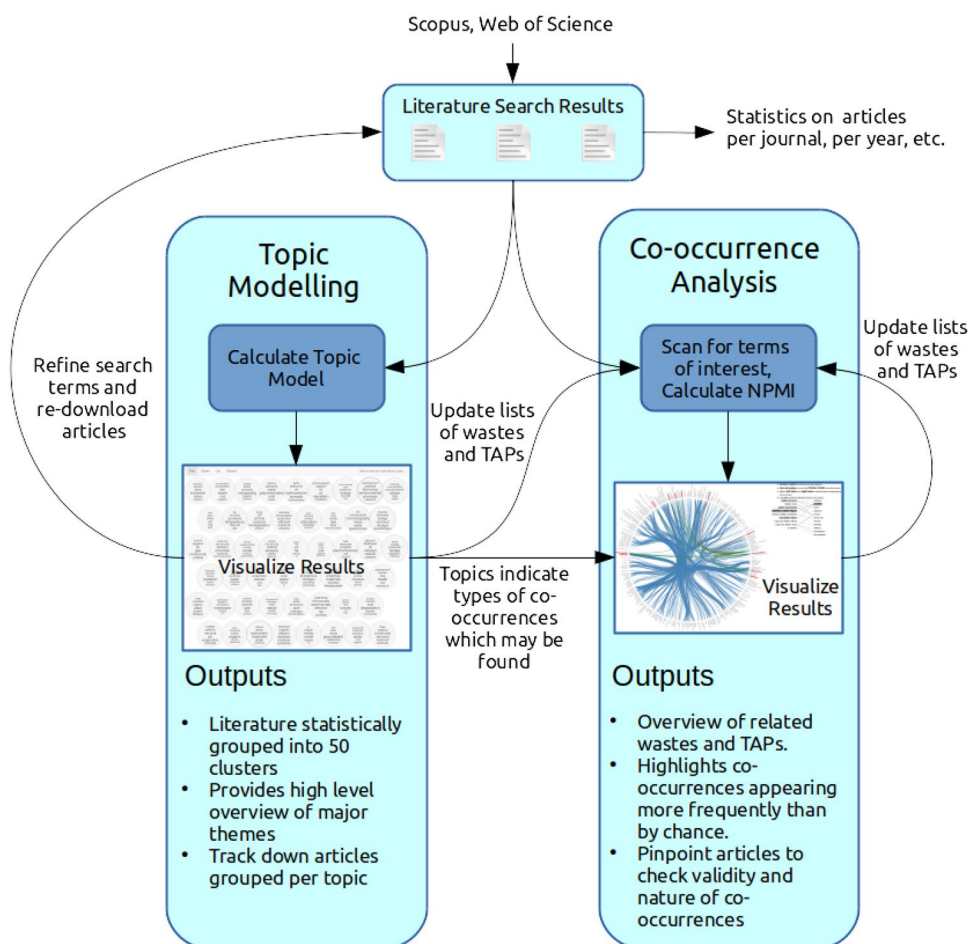
Since the co-occurrence analysis results are extensive and difficult to visualize within a single image, we have also created a series of matrices, introduced in “[Co-occurrence Analysis Results](#)”, showing the NPMI values between TAPs (on rows) and wastes (on columns). As an additional step, we have also applied hierarchical clustering to the rows and columns in these matrices. This clustering serves to re-order the matrix so that rows and columns with similar values are located near each other. In practice, this gives an indication of similarity of wastes and TAPs. For wastes, this clustering may group together materials with similar properties that could serve as substitutes in a particular value pathway. For TAPs, this clustering can indicate chemicals which are related to processes, or relations between products and processes (such as syngas and gasification).

### Iterative Support Between Topic Modelling and Co-occurrence Analysis

These approaches are not meant to operate in isolation. As illustrated in Fig. 3, we used the topic modelling results as a means of quickly evaluating if we had overtly irrelevant literature results, which meant that we had to adjust our Web of Science and Scopus search terms and re-download the results. The results of the topic modelling also gave insight into the types of wastes and TAPs that we could expect to see in the co-occurrence analysis. If we failed to see these wastes or TAPs, then this was an indication that the lists we used to scan the literature for terms was not complete enough and that they should then be complemented.

<sup>7</sup> <https://bl.ocks.org/mbostock/7607999>.

**Fig. 3** Meta-overview of linkages between topic modelling and co-occurrence analysis. The images shown for the topic modelling and co-occurrence analysis are shown in more detail in Figs. 9 and 11 respectively



## Results

In this section we give an overview of the results of the literature collection, the topic modelling and the co-occurrence analysis. More detailed data and results of the topic modelling and co-occurrence analysis are made available online at <http://github.org/isdata-org/mapping-the-bioeconomy>. The static images included below give only a limited impression of the nature of the results. We *highly recommend* that readers view the resources we have placed online in order to get a much richer grasp of the value that is provided by the methods described.

### Literature Collection Results

Through the application of the literature collection methodology, the title, abstract, and keywords from 53,292 articles were collected. Figure 4 gives an overview of the top 20 journals in terms of the number of articles, with a breakdown showing the number of articles found in Scopus and Web of Science, along with the number of duplicate articles found in both. Figure 5 shows the upswing in total articles per year in the collection. Figures 6 and 7 show the

breakdown of the number of journals, books, and conference proceeding collections (6436 total unique), along with the overlap in articles (53,292 total unique).

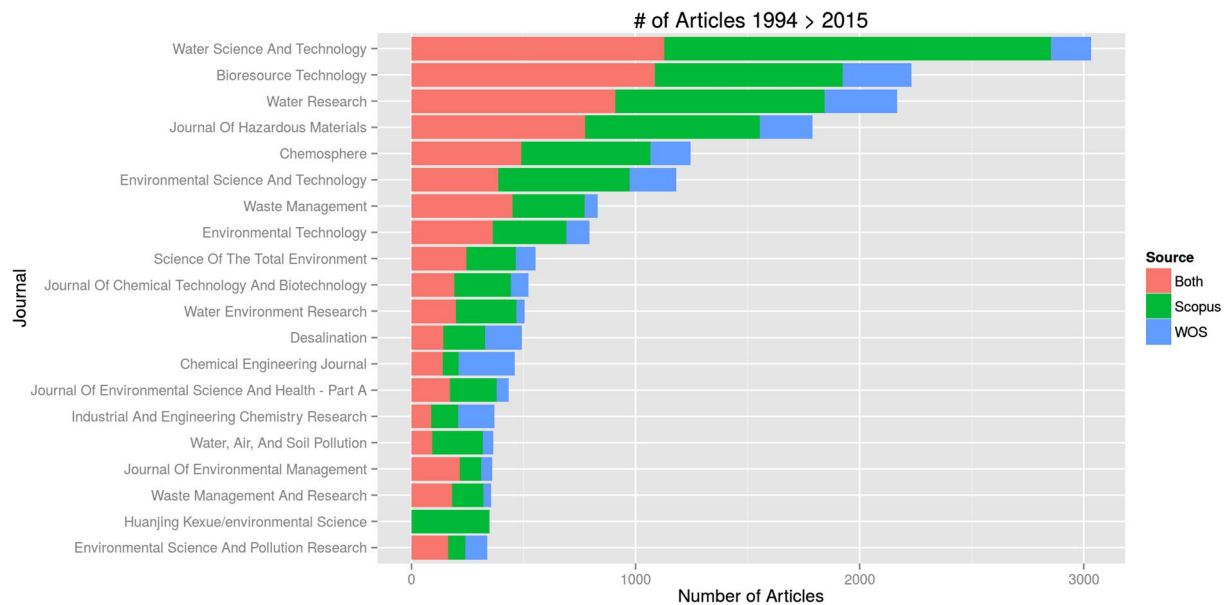
### Topic Modelling Results

The results of the topic modelling method give an overview of the various areas of focus found in the literature collection. Each topic is the result of an automated statistical analysis of collections of words that commonly appear together. Figure 8 presents two of the generated topics in the form of word clouds, where the size of the word represents the weight, or importance, of that word in that particular topic.

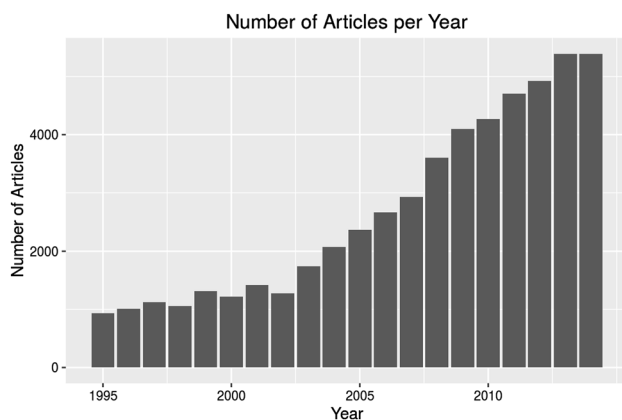
The word cloud on the left indicates that in the literature collected, there are numerous abstracts discussing biodiesel and methanol. The word cloud on the right indicates that we have also collected literature related to fibers and polymers, and that this topic also seems to be discussing the properties of these.<sup>8</sup>

<sup>8</sup> Supporting literature for the Biodiesel topic can be seen at <http://isdata-org.github.io/mapping-the-bioeconomy/TopicModelling/index.html#/topic/25>, with literature for the Polymers/Fibers topic available at <http://isdata-org.github.io/mapping-the-bioeconomy/TopicModelling/index.html#/topic/32>.

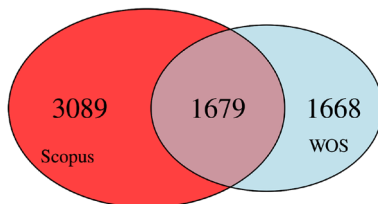




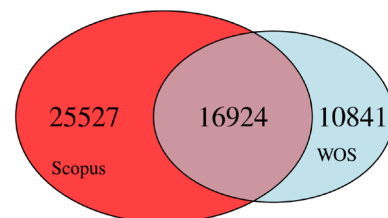
**Fig. 4** Top 20 journals including articles in the study's abstract collection



**Fig. 5** The development of the total number of articles per year in all journals (duplicates removed)



**Fig. 6** Number of journals and collected proceedings in the article collection from Scopus, Web of Science (WOS) and both databases



**Fig. 7** Number of articles collected from Scopus, Web of Science (WOS) and both databases

In Fig. 9, each circle represents a topic and the top six words for that topic are shown.<sup>9</sup> Exploring the complete interactive result of the topic modelling allows for a more in-depth overview of each topic, e.g. showing percent of publications belonging to the topic per year and providing links to the included abstracts. See Fig. 10 for a static view of a specific topic and see the visualization online<sup>10</sup> for the complete interactive topic modelling results.

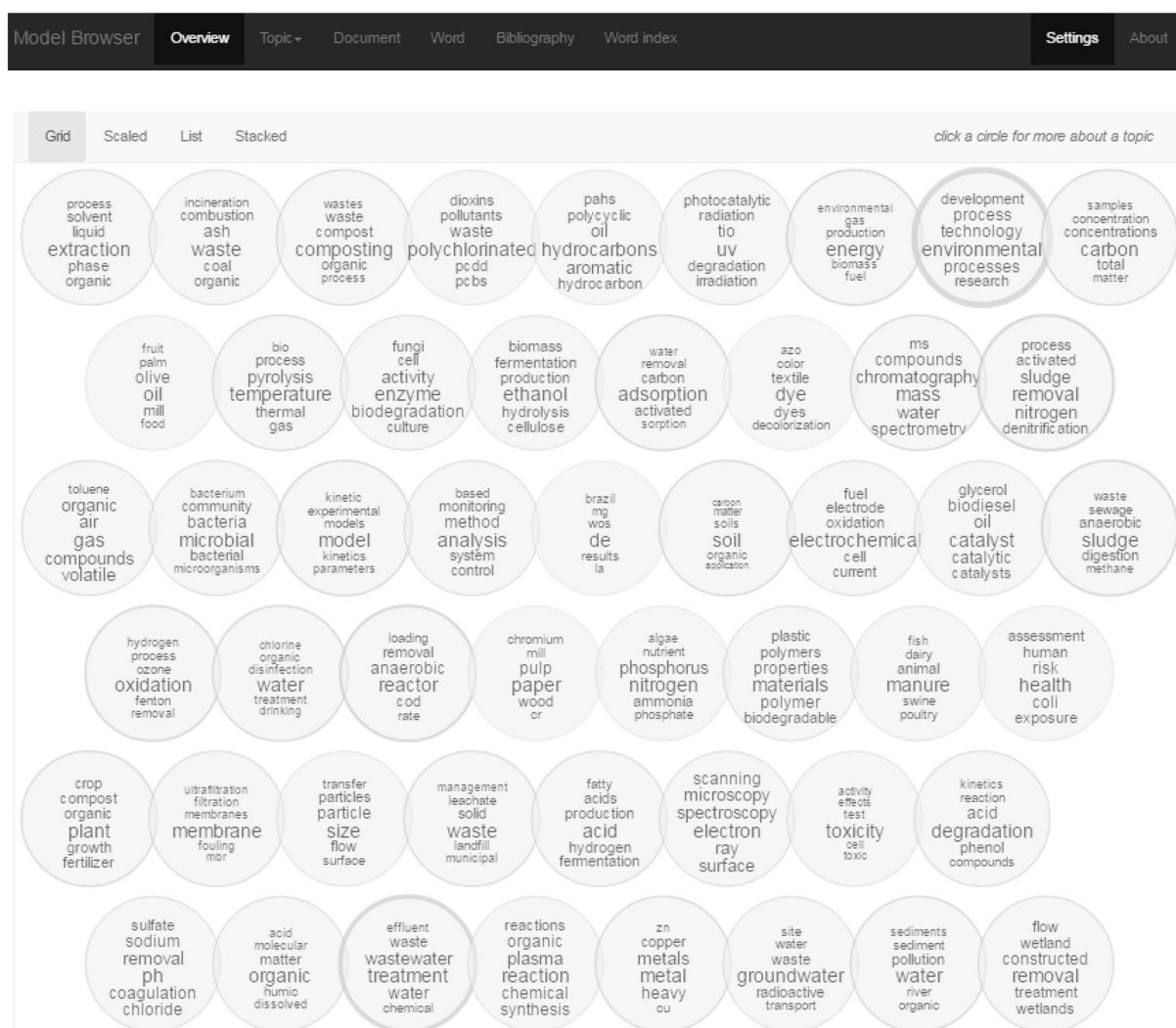
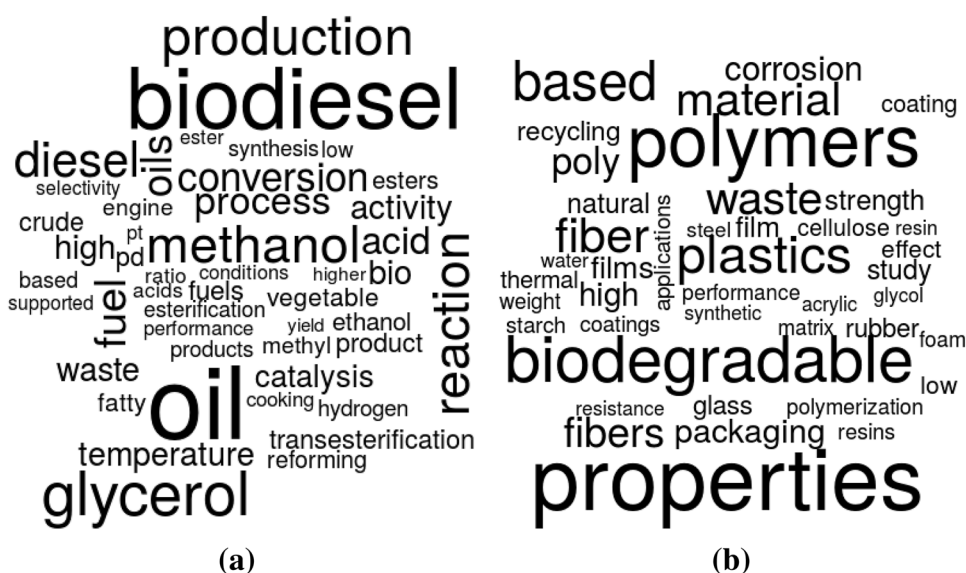
While these topics should be familiar to researchers with experience in the bio-economy domain, the added value of this approach, which we will show later, is that we can drill-down within these topics to show the actual articles that are most representative of the topic.

Many of the topic clusters center on specific waste materials (such as manure or sludge), industries (pulp/paper,

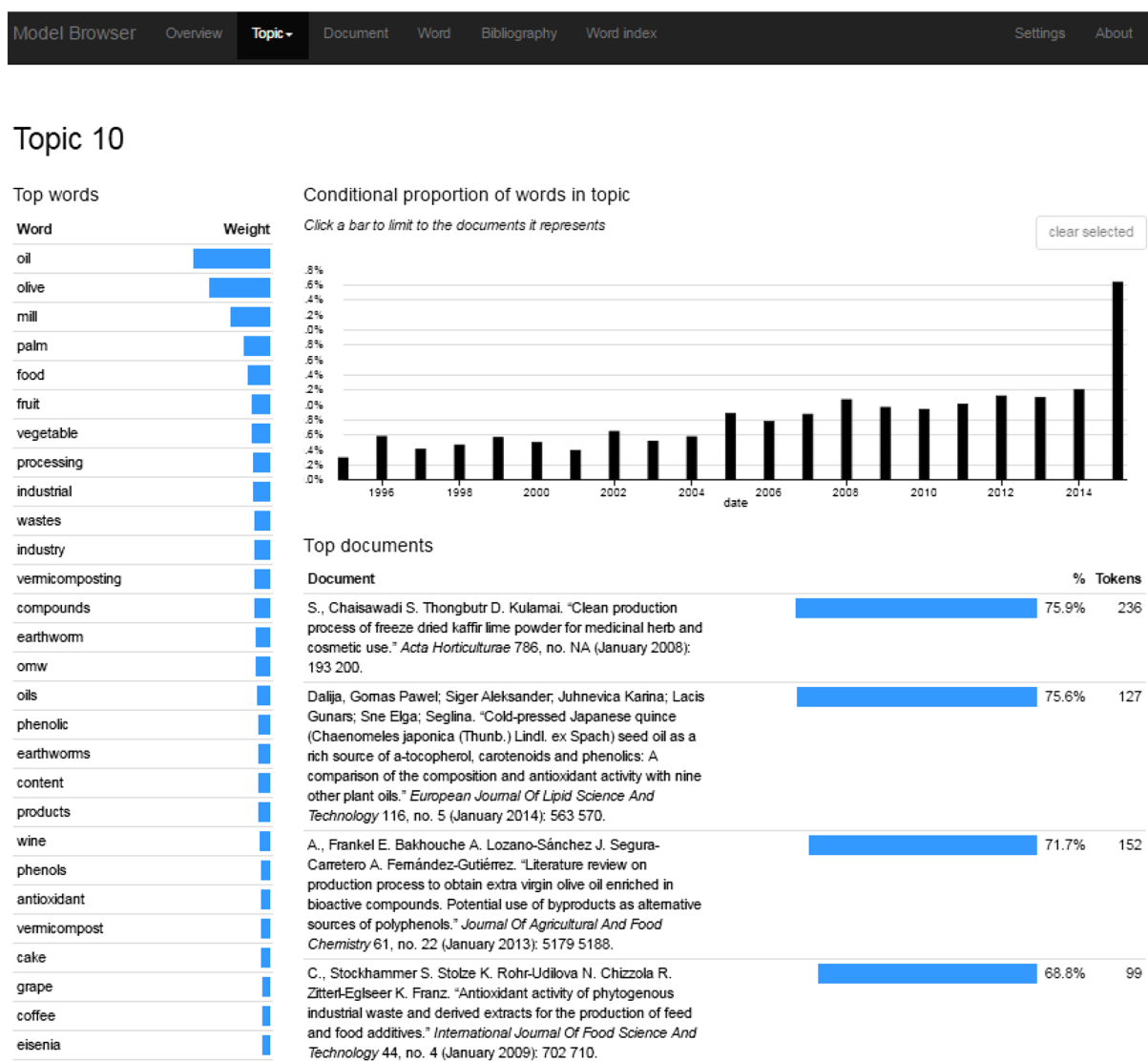
<sup>9</sup> Note that Fig. 8 is showing the same as Fig. 9, except it highlights only two topics and shows more of their words.

<sup>10</sup> <http://is.gd/topicmodel>.

**Fig. 8** Two of the generated topics in the form of word clouds



**Fig. 9** A static result of the topic modelling overview presented in grid format. The words within the *topic circles* illustrate the mix and importance of words within each respective topic. This can be viewed online at <http://isdata-org.github.io/mapping-the-bioeconomy/TopicModelling>



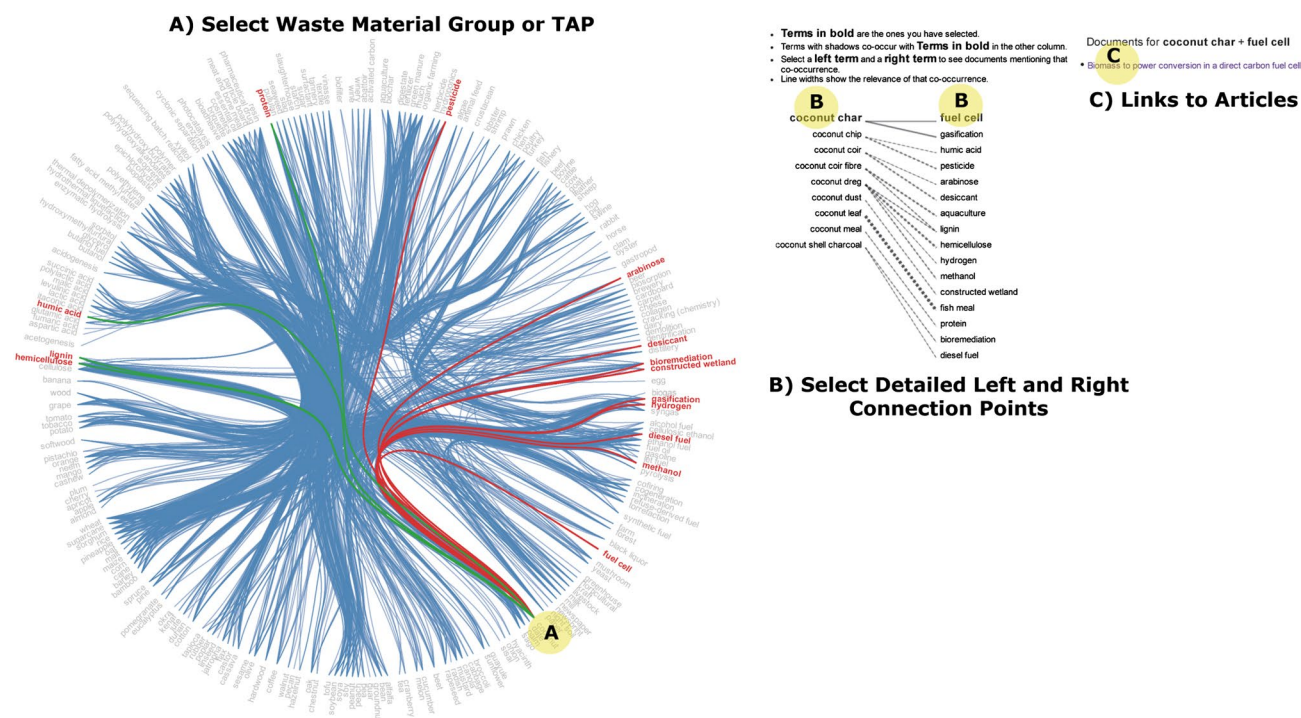
**Fig. 10** A static view of Topic 10 from the topic modelling. Note that the assigned topic numbers may change when the topic modelling analysis is rerun, due to the nature of the algorithm used. This

can be viewed online at <http://isdata-org.github.io/mapping-the-bio-economy/TopicModelling>

textiles, farming), new products (such as ethanol, compost and plant oils), treatment and valuation processes (such as digestion, incineration, microbial fuel cells, constructed wetlands and other various water treatments), or scientific methodologies and measures (such as simulation models, human toxicity, systems analysis, etc.). Therefore, in mirroring the extent of scientific investigations, the topic modelling shows us that our collection of documents addresses an area much broader than straightforward valuation pathways for undervalued resources. However, in reviewing the individual topics and bibliography of the literature collection, a good section of the documents are addressing the area we set out to capture. For example, looking more closely at top articles within Topic 10 (Fig. 10), one can see

that the first two articles deal predominantly with primary raw material valuation, while the third and fourth articles are addressing valuation pathways for secondary materials.

While the topic modelling gives a broad overview of the collected literature (Objective B) and allows for structured explorations of this literature, it does not make the task of identifying a broad range of secondary material valuation pathways (Objective C) much simpler than manually reviewing the full bibliography. Topic modelling only provides a high level overview as it determines topics based on statistical regularities. It is not as useful in locating infrequent patterns representing novel information that may be of interest to researchers. In the next section, the results



**Fig. 11** Example visualization of co-occurrence analysis results. The full interactive version can be viewed online at <https://is.gd/wastecircle>. Section **a** shows a high-level overview of co-occurrences between a wide range of wastes and TAPs. Section **b** shows a more detailed

version of what the user has highlighted in section **a**, and section **c** shows supporting literature for a specific selected combination of waste and TAP

of the co-occurrence analysis take us further in the work toward Objective C.

### Co-occurrence Analysis Results

The co-occurrence analysis provides more detailed insight into the value pathways for secondary organic material from the literature collection. One limitation to be aware of with this approach is that it only examines the co-occurrence of terms in a document, and does not extract information on the nature of the relationship between terms. Because of this it does not effectively differentiate between input materials, process (helper) materials, and output materials. Therefore the material's place in the value chain has to be either inferred or identified by looking into the source literature mentioning the co-occurrence. This limitation can also be of value since it can identify combinations that researchers may not immediately think of (e.g. fish bones or oyster shells as catalysts for biodiesel production).

Figure 11 gives an overview of the graphical user interface (GUI) for the co-occurrence analysis results. The full GUI can be explored at <https://is.gd/wastecircle>. The connections identified through the methodology are summarized in the large circle to the left. Here, waste materials have been grouped according to the waste source.

For example, coffee-bean residue, -bean skin, -fruit peel, -ground, -hull and -husk have all been grouped under coffee.

By placing the mouse pointer over one of the terms (the example in Point A of Fig. 11 is “coconut”), a list of more specific wastes and a list of co-occurring TAPs is presented (points B in Fig. 11). Here co-occurrences between wastes and TAPs are represented by dashed lines, where the width of the line represents the NPMI of the co-occurrence, thus giving an indication of how significant it may be. As shown, some wastes may share the same TAP and vice versa. In the Fig. 11 example, waste materials detailed under the first column of this section (B) include coconut-char, -chip, -coir, -fiber, -dreg, -dust, -leaf, -meal, and -shell. Fuel cell, gasification, humic acid, pesticide, arabinox, aquaculture, diesel fuel, and fish meal are among the many TAPs listed to the right of the coconut waste list. By selecting one specific waste (coconut char in Fig. 11) and one specific TAP (fuel cell in Fig. 11) a list of documents from the literature collection including these co-occurrences is given on the far right (point C). In the example shown, an article by Munnings et al. [20] from the International Journal of Hydrogen Energy is linked.

In exploring the circle on the left, TAPs have been roughly clustered together. For example liquid fuels from



near the bottom of the value pyramid such as alcohol fuel, cellulosic ethanol, diesel fuel, ethanol fuel, fuel oil, gasoline, jet fuel, and methanol have been clustered. It should be noted that neither the TAP nor the waste clustering is perfect at this point, and while the clustering adds a level of structure for the GUI they should be viewed critically and not as definite.

As with waste materials, a user can select a TAP in the circle to explore co-occurrences with that TAP. For example, selecting succinic acid (a precursor to polymers, resins, and solvents) in the circle will give a list of wastes co-occurring with succinic acid in the literature collection. As no synonym database for TAPs is currently applied in the methodology only one TAP will be listed on the left column (point B). When selecting the TAP and the waste material of interest in the columns, one will again be presented with a list of articles documenting these co-occurrences in the literature collection.

The co-occurrence GUI can present a user with value pathways that are surprising, which the user may not have known to look for, such as:

- Producing PHA plastic from milk whey [21],
- The potential to co-digest paper pulp waste and algae sludge [22],
- The use of date palm leaf as a nitrate removing biofilter [23],
- The production of polylactic acid (PLA) from waste textile (jute) fibers [24],
- Oyster shells as a catalyst in biodiesel production [25],
- Bioplastic production from brewery waste [26].

At the time of publishing, 473 specific wastes, and 228 TAPs are presented in the GUI. A total of 2421 unique co-occurrences are represented based on a cutoff value of  $\text{NPMI} \geq 0.2$ . As would be expected, a review article has the most co-occurrences (24) and discusses applications of sugarcane bagasse [27]. A more comprehensive list of terms to scan for (e.g. including chemicals and organisms found via DBpedia Spotlight) would result in far more abstracts ending up in the final analysis.

As mentioned in "Co-occurrence Analysis and Visualization", the co-occurrence analysis results are also presented in the form of matrices showing the NPMI values between combinations of wastes and TAPs. One such example is shown in Fig. 12 for the example of poultry related wastes. Matrices for the rest of the wastes and TAPs, in addition to a link to the raw data in a spreadsheet, are available in Appendix A.6.

In Fig. 12, "chicken litter" and "poultry bedding" are clustered next to each other, and these are indeed synonyms that were not identified as such in our original list of wastes. Some of the results for chicken feathers reflect

those shown in a review of valorization options for keratinous materials [28], which discusses how feathers can be used in fertilizer, animal feed, and in composites.

## Discussion

The discussion is structured according to the procedural methodology. For each method (literature collection, topic modelling, and co-occurrence analysis) we begin by lending a discerning eye to the results of the respective method, then we turn our attention to the methods themselves. Subsequently, we take a step back to analyze the process as a whole in the context of encouraging the growth of the bio-based economy (particularly with our focus on secondary organic resources), before addressing potentials for further research and activities in the area.

### Discussion of the Literature Collection Results

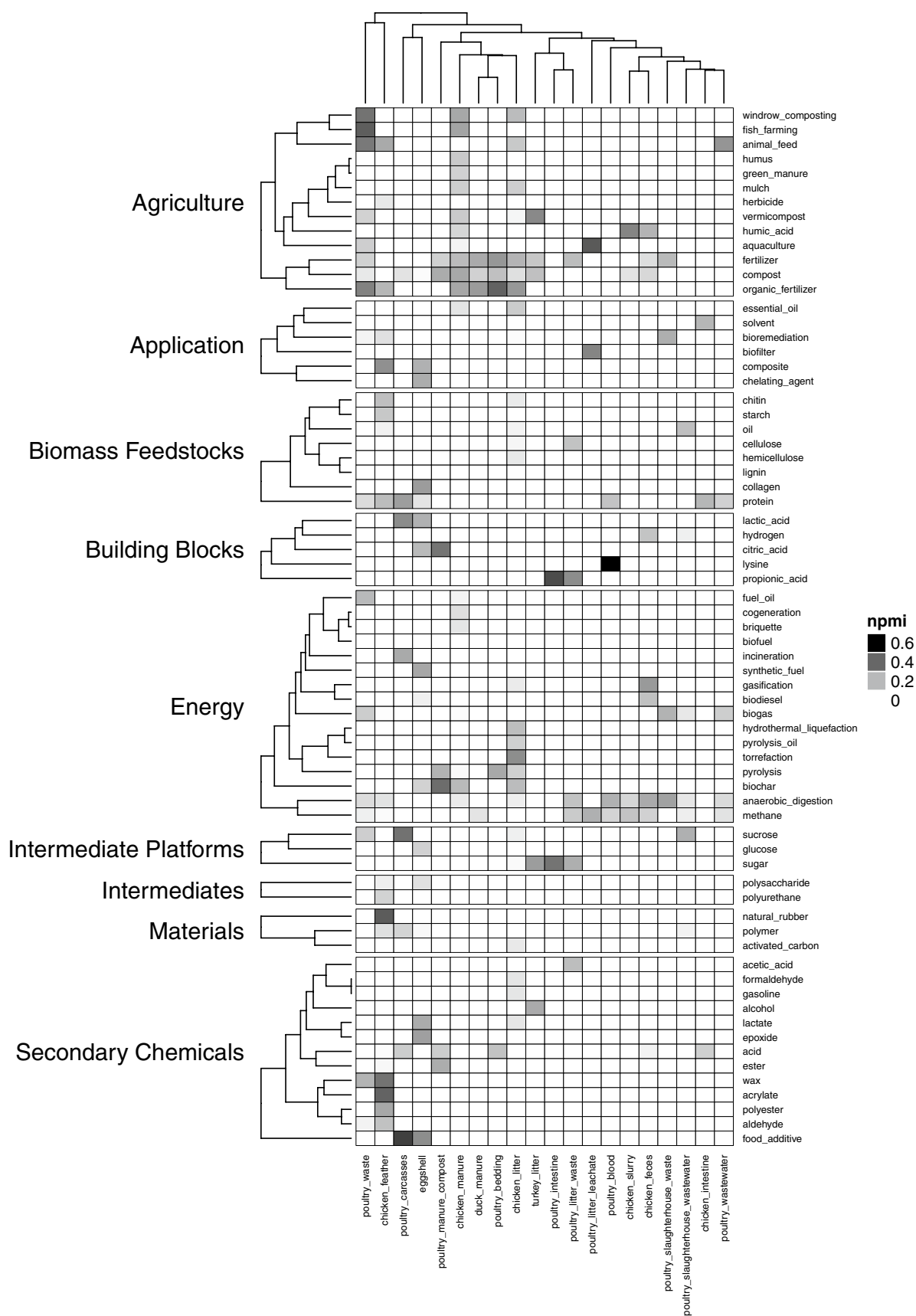
While topic modelling and co-occurrence analysis were used to review and analyze the literature collection in detail, the literature collection in itself presented interesting results. This study shows that there has been a lot of academic activity in the focus area during the last two decades. The amount of literature published in the study is steadily increasing from year to year (especially since 2002), as is the case in many research areas [29].

Looking at Fig. 4, it is notable that research areas outside of traditional waste management are leading in collected article count. Journals such as *Water Science and Technology*, *Bioresource Technology*, and *Water Research* lead in total number of articles<sup>11</sup> (of the total 6436 journals and proceedings). Journals focusing on wastes and by-products specifically, e.g. *Waste Management and Waste Management and Research*, are positioned lower in the rankings with 3–9 times fewer articles collected than the six leading sources. The reasoning behind these numbers can be attributed to in part to 1) the total amount of literature a journal has published, and 2) the specific methods that databases such as Scopus and Web of Science use to retrieve their search results.

Addressing the first issue, additional analysis of the source journals is presented in Table 1. It can be seen that the top six journals had more total articles in our 20 year period (from ca 11,000 to 24,000 total articles), while *Waste Management and Waste Management and Research* had fewer total articles (ca 3700 and 1800 respectively).

<sup>11</sup> Total articles = total number of articles published by a journal from 1995 through 2014, including those not collected in our focused search.





**Fig. 12** Co-occurrence results for poultry

**Table 1** Number of articles collected in our literature search versus total number of articles published by a certain journal in 1995 through 2014

Source	Articles collected via search terms	Total articles published 1995–2014	% of total (%)
Water Science and Technology	2852	14,891	19
Bioresource Technology	1925	13,531	14
Water Research	1843	10,563	17
Journal of Hazardous Materials	1554	13,156	12
Chemosphere	1067	14,720	7
Environmental Science and Technology	974	23,792	4
Waste Management	772	3715	21
Environmental Technology	689	2595	27
Water Environment Research	467	2733	17
Science of the Total Environment	465	12,134	4
Journal of Chemical Technology and Biotechnology	443	3954	11
Journal of Environmental Science and Health—Part A Toxic/Hazardous Substances and Environmental Engineering	380	3266	12
Huanjing Kexue/Environmental Science	349	6361	5
Desalination	328	9284	4
Waste Management and Research	319	1759	18
Water, Air, and Soil Pollution	318	5796	5
Journal of Environmental Management	312	4474	7
Advanced Materials Research	309	132,900	0
Environmental Science and Pollution Research	241	4211	6
Journal of Environmental Quality	235	4624	5

However, Waste Management and Waste Management and Research, are second and fourth of the top 20 journals when ranked by percentage of the journals' total articles included in our search results.

Addressing the second issue, it was found that while Web of Science searches for exact matches to search terms in the title, abstract, and keywords; Scopus will return matches that are part of a word in the title, abstract, keywords, and indexed keywords. Upon deeper investigation, this resulted in words such as “wastewater” matching the search term “waste” in our search string. This further explains the prominence of journals such as Water Science and Technology and Water Research in the collection. However, via topic modelling, it was concluded that the inclusion of wastewater research did not adversely impact the source material to be analyzed via co-occurrence analysis. Indeed much of this material is still inside our broad target area of valuation and treatment of organic wastes and by-products.

### Discussion of the Literature Collection Method

The method used to collect literature from Web of Science and Scopus was useful in gathering large amounts of research literature in a relatively short amount of time. In

future studies it should prove beneficial to consider other collection methods and expanded literature sources.

Other collection methods could improve the depth and breadth of literature surveyed as well as simplify the collection approach. In this study, tools available to typical university personnel were employed. It was later uncovered that some university libraries have more sophisticated connections to Web of Science, Scopus, ScienceDirect, etc. (such as full database downloads), that would allow for collecting more than 500 abstracts per batch (2000 per year) as well as allowing for batch downloading of full documents (not just title, keywords, authors, and abstract). Making use of such resources would exponentially increase the amount of information available to analyze via topic modelling and co-occurrence analysis methods.

Additional literature sources outside academic databases could also complement a meta-study of potential valuation processes and materials. The inclusion of industry journals such as Waste Management World or food and agribusiness trade journals could balance out the gap between research and industry application and innovation. Other sources such as grey literature search engines<sup>12</sup>, patent databases or online case study repositories could also broaden the base of such meta-studies.

## Discussion of the Topic Modelling Results

In this study, topic modelling enabled the authors to fulfill one of the main objectives by identifying clustered topic areas related to organic material resource management. Additionally, topic modelling has shown itself to be a useful tool in checking the quality of the literature collection and co-occurrence methodology results via presenting what potentially unwanted material has been gathered in the literature collection, and by helping to identify what potential connections might be missing in the co-occurrence analysis.

To gain a more clear understanding of the use of these results, one can take the perspective of several actors. A student entering a new interdisciplinary field, such as the bio-based economy, could find these results useful in identifying the multitude of research approaches taken in various corners of the field (e.g. environmental assessments, policy studies, technological optimization studies, innovation studies, environmental management, etc.). Such a student might also identify interconnected fields or processes on the outer boundaries of the system they are studying. The lack of a particular topic may also indicate a research gap. A business analyst for a waste management organization looking to increase their activities in the bio-based economy might use such clustered topic information to map out potential business areas alongside the more common areas (such as waste electronics, plastics, paper, metal, and aggregate recycling). A journal editor in the field could use these results to compare their headline sections with the topics identified across thousands of journals. This might help in identifying areas for new niche sections, or special issues. While this approach can be useful in itself, it is not a replacement for traditional overview methods such as manual macro studies, industry journals, or conference participation. However, we do believe that it can play a role in helping to augment these methods.

## Discussion of the Topic Modelling Method

The approach presented in this article is not the only automated means of visualizing the contents of a large body of literature. There are integrated analysis tools enabling summaries of large result lists from broad search queries in both Web of Science and Scopus. However, these tools are limited to general factors such as journal name, broad subject categories, country, document type, etc. Other analysis tools, such as Thomson Innovation's "Patent Themescape" [30], can perform clustering analysis similar to that

of the topic modelling performed in this study. However, such integrated topic modelling is currently limited to a few exclusive services such as Thomson Innovation's patent database.

## Discussion of the Co-occurrence Analysis Results

The results showed us that insight that can be extracted from the co-occurrence GUI on macro and more detailed levels. On a macro level, we can see that there is an extremely vast amount of research being performed around the valuation pathways for bio-based waste materials. Several levels of the bio-based economy value pyramid are represented in the visualization circle, from low value/high throughput TAPs such as incineration, to higher value/lower throughput TAPs such as cosmetics production.

On the specific waste and TAP levels, one can see a very broad range of resources and processes with co-occur, with an indication of the novelty of this combination. For someone with a particular type of process, this gives a survey of what types of feedstocks could be interesting to explore. For someone with a particular type of waste, this shows which downstream processes may be viable valuation opportunities.

To seek broader feedback on the results of the co-occurrence analysis, the web tool was introduced to 40 undergraduate students at Linköping University studying Industrial Symbiosis. In general, the students found the visualization surprising given the broader than expected array of value pathways presented. The students noted both in person and in an anonymous questionnaire that such a tool showed potential for supporting their studies (especially for projects around environmental technology), and that they found previously unanticipated value pathways quickly through the tool. Their comments also provided several good suggestions for improving the interaction and usability of the online tool (e.g. simplifying the selection of wastes/TAPS by changing the hover selection to click selection and the addition of a search function<sup>13</sup>). The results of the anonymous questionnaire are provided as supplementary material.<sup>14</sup> This includes responses from the students mentioned, as well as responses from a group of academics from which we solicited feedback.

## Discussion of the Co-occurrence Analysis Method

There are three main issues which limit the current analysis: unknown relations between co-occurring terms, the

<sup>12</sup> Grey literature refers to research and reports published outside of conventional academic and commercial publishing channels.

<sup>13</sup> The circle in the online co-occurrence tool can currently be searched by using CTRL-F only.

<sup>14</sup> Survey results can be found at <https://is.gd/circlesurvey>.

completeness of the lists used to filter for co-occurrences, and biases introduced by the co-occurrence metric used.

The co-occurrence analysis only gives us half the picture as what we ultimately want to know are the relationships between the terms. Ideally we expect the relations to be between inputs, outputs and processes, although this can also reflect more complex aspects of the supply chains. For example, a co-occurrence of cattle and ethanol may appear puzzling at first, until one considers that cattle are fed distillers grains which are a byproduct of ethanol production. At a minimum, the co-occurrence gives us evidence that there exists some type of relation between the entities, although further inspection is needed to characterize it.

As mentioned, the lists we created to scan for terms are not meant to be authoritative, but rather to demonstrate the functioning of the method described, and further work can be done to improve these. An issue with the list of TAPs is that while product classification systems such as the CPA [31] do exist, what is lacking is a system for performing entity recognition on the abstracts that is able to link the entities to the corresponding codes.

With the lists of wastes, our method of searching for permutations of terms found in two lists was in some ways too successful and significant efforts had to be taken on de-duplicating variants of terms. Automated methods for this are available to speed up the process, although manual intervention will still be needed to some extent. Ideally, de-duplication efforts should feed into the creation of a waste thesaurus which would contain the preferred and alternative names for wastes, in addition to being able to manage different levels of specificity in the terms which are found.

While lists help to pull out relevant entities, they also bias the analysis and prevent the discovery of novel combinations that would not be considered likely. For example, in manually reviewing our literature collection we came across an article where waste from the processing of salmon was used in the manufacturing of transistors [32]. Transistor manufacturing is not traditionally thought to be part of the bio-economy, and therefore is not represented as one of the TAPs in the co-occurrence visualization.

Finally, particular metrics used to evaluate the co-occurrence of terms have their own biases. For example, NPMI favors terms that do not occur in the text very often. Terms that frequently co-occur with a wide variety of other terms (e.g. many wastes streams and compost) will have a lower NPMI although they may represent valid combinations. If one simply counts the number of co-occurrences, then that will highlight combinations that are already well known, and it will not be immediately clear if co-occurrences with low frequency are simply happening by chance. Appendix A.5 discusses alternative co-occurrence metrics and links to a spreadsheet where these are used on our data set.

At the point of publishing there are no complete online tools for co-occurrence analysis of bio-based document data available with which to compare. However, similar efforts do exist within the bioinformatics realm, with a notable example being the PubGene project [33] and its Coremine Platform.<sup>15</sup>

### Comparison with Traditional Literature Reviews

The methods demonstrated have several key differences from a “traditional” literature review done by hand. While we do not claim that these can replace a traditional review, we believe that they can certainly augment them. Understanding text can require significant domain knowledge, and this is something that computers still struggle with. Topic modelling and co-occurrence analysis are both statistical approaches that help to highlight patterns in the underlying data. There still needs to be a human involved to interpret what these patterns mean, and how, in our case, these patterns can be translated into insights useful for valorization of waste.

As demonstrated, these techniques can quickly process much larger amounts of text than would be feasible by a person. While we did have to spend significant time on the creation of the lists of wastes and TAPs, this was a one-time expense and these lists are available for re-use on new sets of literature.

We would argue that these approaches produce a less biased overview than could be done by a human, who would possess their own interests, perspectives, and background knowledge that would influence their interpretation of the literature. One could counter-argue that the co-occurrence analysis is biased based on the lists that we use, however these help us to consistently scan for mentions of items in literature, and these lists could be extended as well based on expert feedback.

The co-occurrence analysis, especially when presented in matrices that are hierarchically clustered, helps us to also to compare which waste streams (or TAPs) have similar properties or applications. Reading across the rows or columns yields indicative profiles for wastes and TAPs, and these profiles are often automatically built up from the numerous literature sources used.

Finally, the co-occurrence analysis helps with the analysis of large numbers of potential combinations of wastes and TAPs. The analysis of such a large permutation space is extremely difficult with literature reviews done by hand.

<sup>15</sup> <http://www.pubgene.com/products/coremine-platform/>.

## Discussion of the Overall Approach

There are other methods to achieve the overall aims of this study in regards to mapping out pathways for secondary materials in the bio-based economy. Other such methods include keeping up with industry journals and newsletters, reviewing grouped subjects in online encyclopedias, or performing manual literature reviews of a more select set of documents such as case studies and macro studies. One may also conceive other data driven methodologies using, for example, life cycle inventories along with national input output data.

As was shown in the results of this study, there are hundreds of resources to be accounted for in the bio-based economy varying in value and magnitude. Additionally, there are hundreds of processes available to create value from these resources. The approach presented in this article was able to give a very broad overview of these resources and value paths within a limited time, and can be automatically repeated as new literature becomes available. The process can be adapted to several taxonomies of materials and applications. The results arguably give one of the broadest overviews of its kind.

## Further Research

There are many potential further activities in the area of modelling and mapping large research areas via the methodologies presented. Further work with other taxonomies, such as product [31], waste, or industry [34] classifications, would allow for more structured analysis of the materials and industrial sectors involved. Further testing of the use of the outputs in various contexts (for example with a waste management organization, a university student, journal editors, or regional resource networks) would help to verify and improve the methodologies and their intended use cases. Finally, developing the methods to extract information such as environmental consequences, economic data, and volume magnitudes could create added benefits for certain users.

## Conclusion

This study aimed to present methods to deal with increasingly information dense research fields as well as to take steps in supporting the growth of bio-based economies via a wide mapping of innovative value pathways for organic wastes and resources. The approach taken was a multi-method process including the collection of a large body of select academic literature, revealing the clusters of topics included in this literature, and running a detailed co-occurrence analysis to locate the intersections of various wastes

and TAPs (technologies, applications, and products). As a result, proof of concept approaches for clustering topic areas of focus and mapping the co-occurrence of 473 organic waste materials to a wide range of 228 TAPs were described and analyzed in regard to their effect. The results of the analyses gave a significantly wider overview of the many value path potentials for secondary organic resources being researched in the area than previous meta-studies. The methods have proven to be of interest in their applications; however improvements could be made via broader source data, the use of better material and waste taxonomies, and tools for exploring specific results in more detail. The usefulness of results depends on the intended application of the mappings, and is seen as especially interesting for integrated systems engineers, academics with various interests, and perhaps regional authorities.

**Acknowledgements** This work has been funded in part by the Ragnar Sellberg's Foundation.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. European Commission. A Roadmap for moving to a competitive low carbon economy in 2050. Technical report (2011)
2. European Commission. Innovating for sustainable growth: a bio-economy for Europe. Technical report, Publications Office of the European Union (2012)
3. Staffas, L., Gustavsson, M., McCormick, K.: Strategies and policies for the bioeconomy and bio-based economy: an analysis of official national approaches. *Sustainability* **5**(6), 2751–2769 (2013)
4. Eickhout, B.: A strategy for a bio-based economy. Technical report (2012)
5. BioBased Economy Netherlands. BioBased Economy—Bio-raffinage. [http://www.biobasedeconomy.nl/wat-is-biobased-economy/themas/bioraffinage\\_v2/](http://www.biobasedeconomy.nl/wat-is-biobased-economy/themas/bioraffinage_v2/) (2017)
6. Lin, C.S.K., Pfaltzgraff, L.A., Herrero-Davila, L., Mubofu, E.B., Abderrahim, S., Clark, J., Koutinas, A.A., Kopsahelis, N., Stamatelatou, K., Dickson, F., Thankappan, S.: Food waste as a valuable resource for the production of chemicals, materials and fuels. Current situation and global perspective. *Energy & Environmental Science* **6**(2), 426–464 (2013)
7. Mahro, B., Timm, M.: Potential of biowaste from the food industry as a biomass resource. *Engineering in Life Sciences* **7**(5), 457–468 (2007). doi:[10.1002/elsc.200620206](https://doi.org/10.1002/elsc.200620206)
8. Morris, J., Matthews, H.S., Morawski, C.: Review and meta-analysis of 82 studies on end-of-life management methods for source separated organics. *Waste Management* **33**(3), 545–551 (2013). doi:[10.1016/j.wasman.2012.08.004](https://doi.org/10.1016/j.wasman.2012.08.004)
9. Zacharof, M.P., Lovitt, R.W.: Complex effluent streams as a potential source of volatile fatty acids. *Waste and*



- Biomass Valorization **4**(3), 557–581 (2013). doi:[10.1007/s12649-013-9202-6](https://doi.org/10.1007/s12649-013-9202-6)
10. Gajalakshmi, S., Abbasi, S.A.: Solid waste management by composting: state of the art. *Critical Reviews in Environmental Science and Technology* **38**(5), 311–400 (2008)
  11. Korres, N., O’Kiely, P., Benzie, J.A., West, J.: *Bioenergy production by anaerobic digestion: using agricultural biomass and organic wastes*. Routledge, London (2013)
  12. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* **3**, 993–1022 (2003)
  13. Lloyd, S.: Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28**(2), 129–137 (1982)
  14. McCallum, A.K.: MALLET: a machine learning for language toolkit. <http://mallet.cs.umass.edu> (2002)
  15. Werpy, I., Petersen, G.: Top value added chemicals from biomass volume I - results of screening for potential candidates from sugars and synthesis gas top value added chemicals from biomass volume I: results of screening for potential candidates, pp. 76. <http://www.osti.gov/scitech/servlets/purl/15008859-s6ri0N/native/> (2004)
  16. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.B.: Improving efficiency and accuracy in multilingual entity extraction. In: *Proceedings of the 9th International Conference on Semantic Systems*, pp. 121–124, 2013. ISBN 978-1-4503-1972-0. doi:[10.1145/2506182.2506198](https://doi.org/10.1145/2506182.2506198)
  17. Energy Research Centre of the Netherlands (ECN). Phyllis2—database for biomass and waste. <https://www.ecn.nl/phyllis2/> (2012)
  18. Bouma, G.: Normalized ( Pointwise ) Mutual Information in Collocation Extraction. In: *Proceedings of German Society for Computational Linguistics (GSCL 2009)*, pp. 31–40 (2009)
  19. Holten, D.: Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* **12**(5), 741–748 (2006)
  20. Munnings, C., Kulkarni, A., Giddey, S., Badwal, S.P.S.: Biomass to power conversion in a direct carbon fuel cell. *International Journal of Hydrogen Energy* **39**(23), 12377–12385 (2014)
  21. Bosco, F., Chiampo, F.: Production of polyhydroxyalcanoates (PHAs) using milk whey and dairy wastewater activated sludge: production of bioplastics using dairy residues. *Journal of Bioscience and Bioengineering* **109**(4), 418–421 (2010)
  22. Iyovo, G.D., Du, G., Chen, J.: Sustainable bioenergy bioprocessing: biomethane production, digestate as biofertilizer and as supplemental feed in algae cultivation to promote algae biofuel commercialization. *Journal of Microbial & Biochemical Technology* **2**, 100–106 (2010). doi:[10.4172/1948-5948.1000032](https://doi.org/10.4172/1948-5948.1000032)
  23. Hashemi, S.E., Heidarpour, M., Mostafazadeh-Fard, B.: Nitrate removal using different carbon substrates in a laboratory model. *Water Science and Technology* **63**(11), 2700–2706 (2011)
  24. Baheti, V., Mishra, R., Militky, J., Behera, B.K.: Influence of noncellulosic contents on nano scale refinement of waste jute fibers for reinforcement in polylactic acid films. *Fibers and Polymers* **15**(7), 1500–1506 (2014)
  25. Nakatani, N., Takamori, H., Takeda, K., Sakugawa, H.: Transesterification of soybean oil using combusted oyster shell waste as a catalyst. *Bioresource Technology* **100**(3), 1510–1513 (2009)
  26. Wong, A.L., Chua, H., Lo, W.H., Yu, P.H.F.: Synthesis of bioplastics from food industry wastes with activated sludge biomass. *Water Science and Technology* **41**(12), 55–59 (2000)
  27. Loh, Y.R., Sujun, D., Rahman, M.E., Das, C.A.: Sugarcane bagasse—the future composite material: a literature review. *Resources, Conservation and Recycling* **75**, 14–22 (2013)
  28. Chojnacka, K., Górecka, H., Michalak, I., Górecki, H.: A review: valorization of keratinous materials. *Waste and Biomass Valorization* **2**(3), 317–321 (2011)
  29. Bornmann, L., Mutz, R.: Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology* **66**(11), 2215–2222 (2015)
  30. Thomson Innovation. ThemeScape—quick reference guide. [http://info.thomsoninnovation.com/sites/default/files/assets/ti\\_themescape\\_qrg.pdf](http://info.thomsoninnovation.com/sites/default/files/assets/ti_themescape_qrg.pdf)
  31. Eurostat. Glossary: statistical classification of products by activity (CPA) (2013)
  32. Yumusak, C., Singh, T.B., Sariciftci, N.S., Grote, J.G.: Bio-organic field effect transistors based on crosslinked deoxyribonucleic acid (DNA) gate dielectric. *Applied Physics Letters* **95**(26), 341 (2009)
  33. Jenssen, T.K., Lægreid, A., Komorowski, J., Hovig, E.: A literature network of human genes for high-throughput analysis of gene expression. *Nature Genetics* **28**(1), 21–28 (2001)
  34. Eurostat. Glossary: statistical classification of economic activities in the European Community (NACE) (2016)